

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099129 A2

- (51) International Patent Classification⁷: C12Q 1/68
- (21) International Application Number: PCT/GB02/02640
- (22) International Filing Date: 7 June 2002 (07.06.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
0113908.8 7 June 2001 (07.06.2001) GB
- (71) Applicant (*for all designated States except US*): **UNIVERSITY COLLEGE LONDON** [GB/GB]; Gower Street, London WC1E 9BT (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **GRIFFITHS, David, John** [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). **KELLAM, Paul** [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). **WEISS, Robert, Anthony** [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB).
- (74) Agent: **CAMPBELL, Patrick, John, Henry**; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/099129 A2

(54) Title: DESIGNING DEGENERATE PCR PRIMERS

(57) Abstract: A method of designing PCR primers for screening for new members of known virus families.

DESIGNING DEGENERATE PCR PRIMERS

Field of the invention

The invention relates to a method of designing a panel of primers for detecting viruses in a high throughput polymerase chain reaction assay.

Background of the invention

5 All organisms appear to be capable of infection by viruses, including bacteria, animals and plants. Viruses require the use of the cellular translation and transcription machinery to replicate. In the process of replication they often have deleterious effects on the host cell and thus on the host organism. Viruses constitute an important class of pathogens causing many diseases, leading to loss of life in
10 humans and economic loss in the agricultural industries.

Summary of the invention

The polymerase chain reaction (PCR) allows the amplification of a specific region of a polynucleotide. The specificity of the reaction is due to the primers
15 which during the course of PCR bind to the region to be amplified in a sequence specific manner. The invention provides a method of designing primers which can be used in high throughput screening to detect viruses. The method may be used to detect unknown viruses which have not yet been sequenced.

In particular the invention provides a method of designing a panel of
20 degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR, which method comprises

- (a) providing a plurality of amino acid sequences from members of a first
25 virus family,
- (b) comparing the sequences to identify conserved regions,
- (c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region

-2-

identified in (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

5 The invention also provides a method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR, which method comprises

10 (a) providing a plurality of nucleotide sequences from members of a first virus family,

(b) comparing the sequences to identify conserved regions,

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a conserved region identified in (b) and wherein the primer
15 pair is designed to amplify by PCR the nucleotide sequence between the conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

The invention additionally provides a panel of primers which has been designed by the method of the invention.

20

Detailed description of the invention

The invention provides a method of designing a panel of primer pairs which can be used in high throughput virus screening. The method comprises initial steps which deduce the sequences of the primers using computer based calculations, and
25 optional later steps in which the primers are synthesised and tested empirically, for example to determine optimal PCR conditions and/or to select primer pairs with desired further properties.

The panel of primers provided by the method are designed to be capable of detecting unknown viruses based on nucleotide and/or amino acid sequences in the
30 unknown virus which are similar (homologous) to nucleotide and/or amino acid sequences in a known virus. These conserved sequences typically have a role in

-3-

providing a necessary or advantageous activity or property to the virus. Conserved nucleotide sequences may be coding or non-coding sequences.

In one embodiment the conserved sequences code for or are from virus proteins which have the following activities: DNA or RNA polymerase (replicase),
5 topoisomerase (helicase/gyrase), endonuclease (integrase), nucleic acid binding protein, protease, transcription factors, envelope glycoproteins, structural protein (e.g. capsid or nucleocapsid protein).

The panel of primers is designed to detect viruses which are single stranded or double stranded DNA or single stranded or double stranded RNA viruses. The
10 viruses are generally capable of infecting prokaryotic or eukaryotic cells, such as bacterial, animal, plant, yeast or fungal cells. Preferably the viruses are mammalian (preferably primate) or avian viruses, such as human, pig, horse, sheep, goat, cow, chicken, turkey or duck viruses.

The viruses are typically from any combination of the following families:
15 Adenoviridae, Arenaviridae, Arteriviridae, Astroviridae, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Deltavirus, Filoviridae, Flaviviridae, Hepadnaviridae, Herpesviridae, Orthomyxoviridae, Papovaviridae, Paramyxoviridae, Parvoviridae, Picornaviridae, Polydnaviridae, Poxviridae, Reoviridae, Retroviridae, Rhabdoviridae, Togaviridae or Bornavirus.

20 The primers of the panel are capable of detecting unknown viruses in a biological sample. Such a sample either originates from a single individual or is a pooled sample from individuals of the same species. Thus the panel of primers detects viruses which infect the same species (from which the sample originates).

Generally in the method at least 15, 30, 50, 100, 200 or more, typically up to
25 a maximum of 300 different primer pairs are designed. The primer pairs designed in the method bind sequence which is conserved across members of a virus family. The panel which is designed in the method may comprise primer pairs that bind sequence which is conserved across substantially all members of the family or across a subset of the members of the family, for example across all members of a subfamily or of a
30 genus. Generally, the primer pairs bind at least 70%, at least 80%, or at least 90% of the known viruses of the family, subfamily or genus. Preferably less than 10, such as

less than 5, primer pairs will be used for the detection of any given family, subfamily or genus in the panel.

The panel of primer pairs is generally capable of detecting viruses from at least 10, 15, 20, 30 or more families, typically up to a maximum of 35 families.

5 The panel of primer pairs may comprise sets of primer pairs which perform a nested PCR reaction. Generally such a set of primer pairs comprises a first and second primer pair. The first primer pair is able to amplify a template nucleotide sequence from a virus to form a PCR product. The second primer pair is able to amplify a nucleotide sequence using the PCR product generated by the first primer
10 pair as a template. The use of nested sets of primer pairs allows increased sensitivity.

In a preferred embodiment each primer pair is specific for a particular virus family, so that it does not detect viruses of other families.

In the method of the invention the plurality of amino acid sequences or nucleotide sequences are provided from different known viruses of the same family.

15 The amino acid sequences or nucleotide sequences will be for the same protein of the different viruses. Typically at least 5, 10, 20, 50, 100 or more sequences are provided. The maximum number of sequences provided will, for example, be 300 sequences.

Each of the sequences which is provided is typically at least 20, 50, 100, 200
20 or more amino acids or nucleotides in length. In general the maximum length of the nucleotide sequences is 1000 nucleotides and the maximum length of the amino acid sequences is 300 amino acids. The sequences may be obtained from a database of sequences, such as GenBank. The sequences may be obtained from a database comprising virus sequences which are organised into homologous protein families
25 (based on sequence similarity relationships).

In a preferred embodiment the sequences are obtained from the VIDA database (described in Alba et al (2001) Nucleic Acids Research 29, 133-136) or the Virus Division of GenBank. The sequences may be provided in the form of a database, preferably in computer-readable form. The sequences are preferably
30 provided in the form of a computer-readable database constructed using programs which identify homologous protein families, such as GeneTableMaker, MKDOM or

PSCBuilder.

The sequences which have been provided are compared to identify conserved regions. Typically such conserved regions will have a length of at least 12 nucleotides, such as at least 15, 21, 27, 36, 99 or more nucleotides (generally up to a maximum length of 200 nucleotides) or at least 4, 5, 7, 10, 25 or more amino acids (generally up to a maximum length of 50 amino acids).

Across the conserved region the virus sequences which are being provided will of course share identity or similarity. Typically the amino acids or nucleotides in at least 50% of the positions in the region will be the same in at least 50 %, 60%, 70%, or 80% of the viruses of the group (i.e. in the family, genus or subfamily).

The algorithm which identifies conserved regions generally uses a multiple sequence alignment method. The method may comprise (a) aligning all pairs of sequences separately to calculate a distance matrix giving the divergence of each pair of sequences, (b) calculating a guide tree from the distance matrix, and (c) aligning the sequences progressively according to the branching order in the guide tree.

A preferred algorithm for the aligning the conserved sequences is CLUSTALW as described in Thompson et al (1994) Nucleic Acids Research 22, 4673-80. Other algorithms that can be used for aligning sequences are MultAlin (Corpet (1988) Nucleic Acids Research 16, 10881-90) or Jalview (Clamp et al (1998) <http://barton.ebi.co.uk>). BLOCKS of conserved regions of amino acids may be extracted from the multiple alignments, typically using the program Blocks Multiple Alignment Processor. Alternatively the entire process of performing multiple alignments and extracting BLOCKS can be performed using BLOCKMAKER (Henikoff and Henikoff (1994) Genomics 19, 97-107).

The output from the alignment and BLOCK extraction set (i.e. the information describing the identified conserved regions) is then entered into the algorithm which designs the primers. Such output is typically in the form of partial sequences which correspond to the conserved regions (BLOCKS). These BLOCKS are input into a primer design algorithm. In one embodiment such an algorithm is CODEHOP.

In the primer design step the conserved regions which are chosen as targets

-6-

for primers preferably comprise few codons with degenerate counterparts, i.e. preferably the sequence has a low redundancy, such as a redundancy of less than 512 fold, 256 fold or 128 fold. Each primer binds in accordance with Watson-Crick base pairing and thus the binding is sequence specific. Each primer will thus be designed to be wholly or partially complementary to the sequence to which it binds.

Each of the primers typically has a length of at least 8 nucleotides, such as at least 10, 12, 15, 20, 30, 40 or more nucleotides (up to a maximum of 50 nucleotides for example). In one embodiment the primer may comprises at least 2, 4 or 6, up to a maximum of 10 for example, inosine bases. Inosine is able to bind to any of the four nucleotides and therefore use of inosine causes a reduction in effective redundancy.

Each primer pair will be designed so that the PCR product generally has a length of at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides (and typically up to a maximum of 5×10^3 nucleotides long).

Each primer is preferably designed so that it anneals to a single site, i.e. the primer will not bind to any other site in the genome of the relevant virus.

Each primer is preferably designed so that it does not exhibit secondary structure, i.e. the nucleotides in the primer will not bind substantially to any other nucleotide in the primer apart from those to which it is covalently linked. In addition preferably each primer is designed so that it does not bind other primers with the same sequence.

In one embodiment the 3' region, and preferably the 3' terminal nucleotide of the primer binds to the target sequence with high affinity, thus preferably this region or nucleotide comprises a G or C.

Generally each primer is designed to have an annealing temperature of from 30 to 65°C, such as 50 to 60°C or 35 to 45°C. In addition each primer pair may be designed to ensure that the two primers do not bind to each other.

The primers are designed by a computer based algorithm. In one embodiment such an algorithm designs primers according to the following rules:

1) A set of blocks is input, where a block is an aligned array of amino acid sequence segments without gaps that represents a highly conserved region of homologous proteins. A weight is provided for each sequence segment, which can be

-7-

increased to favour the contribution of selected sequences in designing the primer. A codon usage table is chosen for the target genome.

2) An amino acid position-specific scoring matrix (PSSM) is computed for each block using the odds ratio method.

5 3) A consensus amino acid residue is selected for each position of the block as the highest scoring amino acid in the matrix.

4) For each position of the block, the most common codon corresponding to the amino acid chosen in step 3 is selected utilizing the user-selected codon usage table. This selection is used for the default 5' consensus clamp in step 8.

10 5) A DNA PSSM is calculated from the amino acid matrix (step 2) and the codon usage table. The DNA matrix has three positions for each position of the amino acid matrix. The score for each amino acid is divided among its codons in proportion to their relative weights from the codon usage table, and the scores for each of the four different nucleotides are combined in each DNA matrix position.

15 Nucleotide positions are treated independently when the scores are combined. As an option, the highest scoring nucleotide residue from each position can replace the most common codons from step 4 that are used in the consensus clamp.

6) The degeneracy is determined at each position of the DNA matrix based on the number of bases found there. As an option, a weight threshold can be specified
20 such that bases that contribute less than a minimum weight are ignored in determining degeneracy.

7) Possible degenerate core regions are identified by scanning the DNA matrix in the 3' to 5' direction. A core region must start on an invariant 3' nucleotide position, have length of 11 or 12 positions ending on a codon boundary, and have a
25 maximum degeneracy of 128 (this is the default setting of CODEHOP). The degeneracy of a region is the product of the number of possible bases in each position.

8) Candidate degenerate core regions are extended by addition of a 5' consensus clamp from step 4 or 5. The length of the clamp is controlled by a melting
30 point temperature calculation (the CODEHOP default is 60°C) and is usually about 20 nucleotides.

-8-

9) Steps 7 and 8 are repeated on the reverse complement of the DNA matrix from step 5 for primers corresponding to the opposite DNA strand.

In one embodiment CODEHOP (Rose et al (1998) Nucleic Acids Research 26, 1628-1635) is used to design the primer pairs. This program uses the above
5 rules.

The primers designed by the algorithm may then be mapped back to the original sequence to choose primer pairs which provide the desired length of PCR product.

The above-described computer based method is repeated until the desired
10 number of primer pairs have been designed. Optionally the primer pairs can then be synthesised and tested. They are typically tested to determine the optimal conditions for using the primers in a PCR reaction.

The PCR reaction is carried out in a PCR mixture that generally comprises the following: the template polynucleotide (which will be amplified in the event of
15 virus detection), one or more primer pairs designed as described above, a polymerase enzyme (typically a DNA polymerase, such as Taq polymerase), deoxynucleotide triphosphates (dATP, dTTP, dCTP and dGTP) and a suitable buffer.

The PCR reaction generally comprises cycles of the following steps: a denaturation step, a primer annealing step and a polynucleotide synthesis step.
20 Typically the PCR reaction comprises at least 25 cycles, such as 30, 35, 40 or more cycles, up to a maximum of 60 cycles for example. Generally in the denaturation step the PCR mixture is heated to a temperature at which the polynucleotides in the PCR mixture (in particular the polynucleotide region to be amplified) denature to single stranded form. The denaturing temperature is generally from 85 to 98°C.

25 In the primer annealing step the primers bind to template nucleotide sequence in a sequence specific manner. This step is generally carried out at a temperature of from 30 to 65°C. In the polynucleotide synthesis step the polymerase replicates/synthesises nucleotide sequence based on template sequence by addition of nucleotides to the 3' end of the bound primers. This step is generally carried out at
30 about 72°C.

In one embodiment the primers are tested for their ability to amplify one or

more of the plurality of nucleotide sequences from known viruses which were used to design the primers, or in the case of amino acid sequences from known viruses being used to design the primers the primers may be tested for their ability to amplify the nucleotide sequence from the virus which encodes the amino acid sequence.

- 5 The primers may be tested in a range of buffer conditions to determine optimal buffer conditions for PCR using the primers. The buffer conditions which may be tested include pH (typically between 7 and 10), magnesium concentration (typically from 0.5 mM to 5 mM), potassium chloride (typically from 0 to 100 mM), ammonium chloride (typically 0 to 100 mM), glycerol (typically 0 to 20%),
10 dimethylsulphoxide (typically 0 to 20%), ethanol (typically 0 to 20%), sorbitol (typically 0 to 20%) or betaine (typically 1M betaine).

The primers may be tested at a range of different temperatures to determine the optimal temperatures in the PCR reaction. Preferably the primers are tested in PCR reaction in which a range of primer annealing temperatures are tested.

- 15 Typically the range of temperatures is from 30 to 65°C.

The panel of primer pairs or a group of primers within the panel may be designed to be used together on the same plate (i.e. using the same thermal cycles). Thus such primer pairs will be designed to work at the same annealing temperature.

- In one embodiment a group of primer pairs within the panel are designed to
20 have similar optimal conditions for use in PCR so that they can be used optimally in the same well or reaction vessel, i.e. that they can be used in multiplex PCR. Such a group typically comprises at least 2, 3, 4, 5, 6 or more primer pairs (up to a maximum of 8 primer pairs for example).

- To provide such primer pairs the computer based method steps may be used
25 to design primer pairs which are calculated to have similar annealing temperatures and/or the primers are tested to select primer pairs which can be used optimally together. Such testing typically determines whether the primers work optimally with the same buffers and/or whether the primers have similar annealing temperatures.

- In one embodiment at least one or both primers of each primer pair in the
30 group carries a label. Typically at least one of the primers in each primer pair will carry a different label from that used for the other primer pairs. The PCR product

generated by labelled primers carries the labels present on the primers. Thus after the group of primers have been used for PCR in the same well detection of the labels in the PCR products can be used to deduce which PCR product was formed from each primer pair. In one embodiment all forward primers of the group are labelled with one colour and the reverse primers are labelled with a different colour.

In a preferred embodiment the primers are labelled with a fluorescent label, such as fluorescein based labels (e.g. fluorescein isothiocyanate). Different primer pairs may be labelled with fluorescent labels of different colours. The fluorescent labels which are used may be capable of detection by a Beckman CEQ2000™ or Applied Biosystems A3700™ fluorescent DNA analyser. The fluorescent labels may be obtained from Beckman Coulter or Applied Biosystems.

Another way of being able to determine which PCR products are generated by which primer pair is for each primer pair in the group to generate a PCR product of different size to the PCR products generated by the other primer pairs of the group. Typically each PCR product which is generated by the group of primers differs in size from all the other PCR products by at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides. Each PCR product may for example differ in size from all other PCR products by up to a maximum of 3000 nucleotides.

The following Example illustrates the invention:

Example

The Example below refers to Figure 1 which shows how primers were designed using a database known as 'VIDA', and computer programs known as 'CLUSTALW', 'BLOCKMAKER' (or 'BLOCKS') and 'CODEHOP'.

Designing a panel of primers

A panel of primers was designed for detecting unknown viruses from the family Herpesviridae according to the strategy shown in Figure 1. The amino acid sequences of herpes virus DNA packaging protein UL15 were obtained from the VIDA database (Alba et al, see above). These sequences are shown in Table 1.

The sequences obtained from the VIDA database were then imported into CLUSTALW. This compares the protein sequences to identify conserved regions

-11-

and then aligns the sequences according to the conserved regions. The alignment produced by CLUSTALW is shown in Table 2.

The BLOCKMAKER program was then used to extract blocks of conserved aligned sequences which do not contain gaps from CLUSTALW and enter them into
5 CODEHOP. The primer sequences were then designed by CODEHOP using the conserved sequences. The output from the CODEHOP program is shown in Table 3. The 'Complement of Block' sequences shown in Table 3 shows the sequence of the other strand allowing primers to be designed for amplification in the opposite direction.

10

Table 1. All protein sequences of DNA packaging protein UL15 extracted from VIDA.
Here written as a list and unaligned.

```

>gi_10180719
5  MFGGLLGEETKRHFERLMKTNDRLGASHRNEIRDGMVDAPFLNFAIPVPRRHQTVMPAIGILHNCC
   DSLGIYSAITTRMLYSSIA CSEFDELRRDSVPRCYPRITNAQAFLSPMMRVANSII FQEYDEMECAAHR
   NAYYSTMNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVNKLDDHTVKKRARIDAPS YDKLHGTLLELFQ
   KMILMHATYFVTSVLLGDHAERAERLLRVAFDTPHFS DIVTRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYTSHIRKAIEPVFEDIGDRLRRWFGAHRVDHVKG ETTITFSFPSGLKSTVTFASSHTNSI
10  RGQDFNLLFVDEANFIRPDVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
   EHMKHVTDYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMQEII GGGVVDRTICQGDRSIFTASA
   IDRFLIYRPSTVNNQDPFSQDLYVYVDPAPTANTKASGTGVA VIGKYGTDYIVFGLEHYFLRALTGESSD
   SIGYCVAQCQLIQICAIHRKRFGVIKIAIEGNSNQDSAVAIA TRIAIEMISYMKA AVAPT PHNVSPYHKS
   NGTDVEYPYFLLQRQKTTAFDFFIAQFN SGRVLASQDLVSTTVSLTDPVEYLT KQLTNISEVVTGPTCT
15  RTFSGKKGGNDDTVVALTMVYISAHIPDMAFAPIRV
   >gi_7673189
   MFGGLLGEETKRHFERLMKTNDRLGASHRNEIRDGMVDAPFLNFAIPVPRRHQTVMPAIGILHNCC
   DSLGIYSAITTRMLYSSIA CSEFDELRRDSVPRCYPRITNAQAFLSPMMRVANSII FQEYDEMECAAHR
   NAYYSTMNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVNKLDDHTVKKRARIDAPS YDKLHGTLLELFQ
20  KMIFDACHLFCNFCFTWRSRRASERLLRVAFDTPHFS DIVTRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYISHIRKAIEPVFEDIGDRLRRWFGAHRVDHVKG ETTITFSFPSGLKSTVTFASSHTNSI
   RGQDFNLLFVDEANFIRPDVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
   EHMKHVTDYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMQEII GGGVVDRTICQGDRSIFTASA
   IDRFLIYRPSTVNNQDPFSQDLYVYVDPAPTANTKASGTGVA VIGKYGTDYIVFGLEHYFLRALTGESSG
25  SIGYCVAQCQLIQICAIHRKRFGVIKIAIEGNSNQDSAVAIA TRIAIEMISYMKA AVAPT PHNVSPYHKS
   NGTDVEYPYFLLQRQKTTAFDFFIAQFN SGRVLASQDLVSTTVSLTDPVEYLT KQLTNISEVVTGPTCT
   RTFSGKKGGNDDTVVALTMVYISAHIPDMAFAPIRV
   >gi_5689285
   MFGGALGESAKKHFERLLRDRNERLGASRKN ECLARGGSLVDAPFLNFAISVPRRHQTVMPAVGTLHDCC
30  DGTGIYSAIATRLLYAGIVSSEFGEVRRESLSNGHISKRNREALLAPTLTRVANSITFHEYDDAQCAAHR
   NAYYSTMTNFTGSMRTSDAFQQLASFIDRF SKLLAASF KDVNILDRNNAPKRARITAPS YDKPHGTLLELFQ
   KMILMHATYFLTSLLEDHAERAERLLRVIFDI PDFSDAATRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYTSHIRKAIEPVFEEIGDRLRRWFGTQCVDHVKG ETTITFSFPSGSRSTVTFASSHTNSI
   RGQDFNLLFVDEANFIRPDVQTIIGFLNQANCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
35  EHMKHVTNYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMKEII GGITMDRNTCQGD RGVTASA
   VERLLLYRPSTVRNQDILSRDLYVYVDPAPTANTRASGTGIA VIGRYGADYIIFGLEHFFLRALTGESAD
   AIGECAAQCIAIQICAIHCERFGTIRVAVEGNSNQDSAVAIA TRISIDLASYVQSGVAPAPHDVCFYH SKP
   AGSNVEYPFFLLQRQKTAADFDFIARFNSGRVLASQDLVSTTISLSTDPVEYLT KQLTNLSEVVTGATGT
   RTFSGKKGGYDDTVVALVMVYISAHASDATFAPIRGVEATCRGPTEA
40  >gi_1869837
   MFGQQLASDVQQYLERLEKQRQQKVGVD EASAGLTLGGDALRVFPLDFATATPKRHQTVVPGVGT LHDCC
   EHSPLFSAVARRLLFNLSLVAQLRGRDFG GDHTAKLEFLA PELVRVARLRFRECAPEDAVPQRNAYYSV
   LNTFQALHRSEAFRQLVHFVRDFAQLL KTSFRASSLAETTGP PPKRAKVDVATHGQTYGTLELFQ KMILM
   HATYFLAAVLLGDHAEQVNTFLRLVFEI PLFSDTAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFR
45  GIKIGYTAHIRKATEPVFDEIDACL RGFSGSRVDHVKG ETTISFSFPDGSRSSTIVFASSHTNGIRGQDF

```

-13-

NLLFVDEANFIRPDVQTIMGFLNQANCKIIFVSSNTNGKASTSFLYNLRGADELINNVVTYICDDHMPR
 VVTHTNATACSCYILNKPVFITMDGAVRRADLFLPDSFMQEIIGGQARETGDDRPVLTSAKERFLLYR
 PSTTTNSGLMAPELYVYVDPFAFTANTRASGTGIAVVGGRYRDDFIIFALEHFFLRALTGSAPADIARCVVH
 SLAQVLALHPGAFRSVRVAVEGNSSQDSAVAIAITHVHTEMHRILASAGANGPGPELLFYHCEPPGAVLY
 5 PFFLLNKQKTPAFEFYIKKFNSGGVMASQELVSVTVRLQTDPEYLLSEQLNLIETVSPNTDVRMYSGKR
 NGAADDLMVAVIMAIYLAAPTGIIPAFFPITRTS
 >gi_59501
 MFGQQLASDVQQLERLEKQRLKVGADSEASAGLTMGGDALRVPFLDFATATPKRHQTVVPGVGTLLHDC
 EHSPLFSAVARRLLFNSLVPAQLKGRDFGDDHTAKLEFLAPELVRAVARLRFKECAPADVVPQRNAYYSV
 10 LNTFQALHRSEAFRLVHFVRDFAQLLKTSFRASSLTETTGPPKKRAKVDVATHGRTYGTLELFQKMILM
 HATYFLAAVLLGDHAEQVNTFLRLVFEIPLFSDAAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFR
 GIKIGYTAHIRKATEPVFEEIDACLRGWFSGARVDHVKGETISFSFPDGSRTIVFASSHNTNGIRGQDF
 NLLFVDEANFIRPDVQTIMGFLNQANCKIIFVSSNTNGKASTSFLYNLRGADELINNVVTYICDDHMPR
 VVTHTNATACSCYILNKPVFITMDGAVRRADLFLADSFQEIIGGQARETGDDRPVLTSAKERFLLYR
 15 PSTTTNSGLMAPDLYVYVDPFAFTANTRASGTGVAVVGGRYRDDYIIFALEHFFLRALTGSAPADIARCVVH
 SLTQVLALHPGAFRGVRVAVEGNSSQDSAVAIAITHVHTEMHRLASEGADAGSGPELLFYHCEPPGSAVL
 YPFFLLNKQKTPAFEFYIKKFNSGGVMASQEIIVSATVRLQTDPEYLLSEQLNLIETVSPNTDVRTYSGK
 RNASDDLMVAVIMAIYLAQAQAGPPHTFAPITRVS
 >gi_2605992
 20 MFGKALSRETIQYFETLRKEVQSRSGAKNRAEAQGTGGEDDVKTAFLNFAIPTPQRHQTVVPGVGTLLHDC
 CETAQIFASVARLLFRSLSKWRGGESKERLDPSSVEAYVDPKVKQALKTISFVEYNDAEARSERNAYYS
 IMNTFDSLRSDDAFHQVANFVARFSRLVDTSFNGADLDGDDGQTSKRIKVDVPTYGKQRTLELFQKMIL
 MHATYFIAAVILGDHADRIAGFLKMFNTPEFSDATIRHFRQRATVFLVPRRHGKTWFLVPLIALALATF
 KGIKIGYTAHIRKATEPVFDEIGARLRQWFGNSPVDHVKGENISFSFPDGSKSTIVFASSHNTNGIRGQD
 25 FNLLFVDEANFIRPEAVQTIIGFLNQTNCKIIFVSSNTNGKASTSFLYNLKGAADELINNVVTYICDEHME
 RVKAHTNATSCSCYILNKPVFITMDGAMRNTAELFLPDSFMQEIIGGNGISGAHRDEPVFTKTAQDRFLL
 YRPSTVANQDIMSNLYVYVDPFAFTNAMASGTGVAVVGGRYRSNWIIFGLEHFFLSALTGSSAELIARCV
 AQCLAKVFAIHSRPFDSVRVAVEGNSSQDAVAIATNIQLELNTLRQADVVMHMPGTVLFYHCTPPGSSVA
 YPFFLLQKQKTGAFDHFIFAFNSGLVLASQELISNTVRLQTDPEYLLTQMKNLTEVITGTSETRVFTGK
 30 RNASDDMLVALVMAYMASLPPTTNAFSSLSLQ
 >gi_330792
 MFGRLVGRETVQYFEALRREVQARRGAKNRAEAQNGGEDDAKTAFLNFAIPTPQRHQTVVPGVGTLLHDC
 CETAQIFASVARLLFRSLSKWQSGEARERLDPASVEAYVDPKVRQALKTISFVEYSDDEARSERNAYYS
 IMNTFDALRSDDAFHQVASFVARFSRLVDTSFNGADLDGDDGQASKRARVDVPTYGKQRTLELFQKMIL
 35 MHATYFIAAVILGDHADRIAGFLKMFNTPEFSDATIRHFRQRATVFLVPRRHGKTWFLVPLIALALATF
 KGIKIGYTAHIRKATEPVFDEIGARLRQWFGNSPVDHVKGENISFSFPDGSKSTIVFASSHNTNGIRGQD
 FNLLFVDEANFIRPEAVQTIIGFLNQTNCKIIFVSSNTNGKASTSFLYNLKGAAADDLLNVVTYICDEHME
 RVKAHTNATACSCYILNKPVFITMDGAMRNTAELFLPDSFMQEIIGGNGVSGAHRDEPVFTKTAQDRFLL
 YRPSTVANQDIMSSDLYVYVDPFAFTNAMASGTGVAVVGGRYRSNWWVFGMEHFFLSALTGSSAELIARCV
 40 AQCLAQVFAIHKRPFDSVRVAVEGNSSQDAVAIATNIQLELNTLRADVVPMPGAVLFYHCTPHGSSVA
 YPFFLLQKQKTGAFDHFIFAFNSGVLASQELVSNVTVRLQTDPEYLLTQMKNLTEVVTGTSETRVFTGK
 RNASDDMLVALVMAYVLSLPPTSDAFSSSLPAQ
 >gi_971317
 MFGGAVGEQSARYFQRLLRERQRRAAERGARDGGGGGEDDARVPFLDFAVAAPKRHQTVVPGVGTLLH
 45 GYCELAPLFAATASRLLLTSMARAEAGLNTGTGEAHVSRELAVLSALRFAAHPFAEAAAHNCNAYHSMVA
 ALESMRASGAFAQVAFAVARFSRLVGTSFSLGGGDDADPPRAKRARVEPPSGQTRGALELFQKMILMPA

-14-

TYFVAATLLGEHAERIGAFRLVAFNTPDFSDAAVAHFRQRATVFLVPRRHGKTWFLVPLIALALATFKGI
 KIGYTAHIRKATEPVFEEIVARLRQWFGGERVDHVKGAVISFSFPD GARSTIVFASSHNTNGIRGQDFNL
 LFDVEANFIRPEAVQTIVGFLNQASCKIIFVSSTNTGKASTSFLYNLKGASDGLLNVTYICNEHTPRVA
 AHGGATACSCYVLNKPVFITMDAAARNTAETFLPNSFMQEIIGGGEVARRAEPAAVFTRAAGEQFLLYRP
 5 STAAARGPWPERLYMYIDPAFTSNARASGSGIAVVGHRGSLVLGLEHFFLPALTGSSAAEIARCAVRC
 FAQVMVHRRRLDGLFVAVEGNSSQDSAVAIALGVRRRLDLSAASGAVPMPAETRFYHCRPPGSAVAYPF
 FLLQKQKTAADFHFIRLFNSGRVVASQDLASLTVRLQTDPEYVLFELQNLTESTAGPGGARAFSGKRRG
 AADDLMVALVMAVFGSLPPTDGAFCPLAPRPPAD
 >gi_5869808
 10 MSLIMFGRITLGEESVRYFERLKRRLDERFGTLESPTPCSTRQGSIGNATQIPFLNFAIDVTRRHQAVIPG
 IGTLHNCCEYIPLFSATARRAMFGAFLSSTGYNCTPNVVLKPWRYSVNANVSPELKKAVSSVQFYEYSPE
 EAAPHRNAYSGVMNTFRAFLSDSFCQLSTFTQRFSLVETSFSFIEECGSHGKRAKVDVPIYGRYKGTLL
 ELFQKMLMHTTHFISVLLGDHADRVDCLRTVFNTSPVSDSVLEHFKQKSTVFLVPRRHGKTWFLVPL
 IALVMATFRGKIVGYTAHIRKATEPVFEGIKSRLEQWFGANYVDHVKGESITFSFTDGSYSTAVFASHN
 15 TNGIRGQDFNLLFVDEANFIRPDAVQTIVGFLNQTNCKIIFVSSTNTGKASTSFLYNLRGSSDQLLNVT
 YVCDHMPRVLASDVTACSCYVLNKPVFITMDGAMRRTADLFMADSFVQEI VGGKQKNSGGVGFDRPLF
 TKTARERFILYRPSTVANCAILSSVLYVYVDPFTSNTRASGTGVAIVGRYKSDWII FGLEHFFLRALTG
 TSSEIGRCVTQCLGHILALHPNTFTNVHVSIEGNSSQDSAVAISLAIAQQFAVLEKGNVLSAPVLLFY
 HSIPPGCSVAYPFFLLQKQKTPAVDYFVKRFNSGNI IASQELVSLTVKLGVDPEYVLCQQLDNLTEVIKG
 20 GMGNLDTKTYTGKGTGTMSDDLMLVALIMSVYIGSSCIPDSVFMPIK
 >gi_5708110
 MLGKESVEIVKRYDALRKRITMERGPDDVDGQEMSDSNFITTASICDRNDSARDTMNSPASRFQFAIDVP
 QRHQACIAPIGSFHNCCAISRAFSYMASEIIYENLASYSTKYTDTDAALNDLQVSPKRQLFTGAEDSIL
 PALRQKLANLNFARFAPSDSLIHDKAFDGINNGYRGFVKSDEFSQLNNFIYRFHTLLKKSFGQASNDYK
 25 RAKLEKTTSEQRDGTLELFQKMLMHATYFASSICLGEGSTERSNRYLSTVFNSTLFSENIIQHFRQRTT
 VFLVPRRHGKTWFLVPLISLLVSSFEGIRIGYTAHLRKATEPVFIEIFTRLYKNFGAKQVEQVKGETITF
 TFRNGNKSIVFASSQNTNGLRGQDFNFLVDEANFIKPAALHTVMGFLNQTNCKLFFVSSTNTCHSNTS
 LLYNLKGKINSLLNVVTYICDEHMPETQKRTDVTTCSCYVLHKPVFVSMDSVRNTADLFVKDSFMHEIA
 GGRAGKYDSRDLTVPRALDQFLIYRPSTSSKPNISGLGKILTVYVDPFTTNRSASGTGIALVTALRDS
 30 MVLMGAEHFYLDALTGEAALEIAQCVYLCIAYCCLIHAGAFREIRIAVEGNSSQDSAAAIAGNLTLELDS
 LRRRLGFSLTFAHSRQPGTAMAHFPYLLNKQKSRAFDLFLVSLFNSGRFMASQELVSNTLVLSKDPCEYLV
 DQIRNITVTHGQGPDSFRFTSGKQGRVPDDMLVAAMSTYLALEGSPTAGYHPIAPIGRQRPA
 >gi_1813970
 MLRGDSAAKIQERYAELQKRKSHPTSCISTAFNTVATLCRKRYQMMHPGLAHS CNEAFLPLMAFCGRH
 35 RDYNSPEESQRELLFHERLKSALDKLTFRPCSEEQRASQKLDALTELYRDPQFQQINNFMDFKKNLDG
 GFSTAVEGDAKAI RLEPFQKNLLIHVIFFI AVTKIPVLANRVLQYLIHAFQIDFLSQTSIDIFKQKATVF
 LVPRRHGKTWFIPIISFLLKHMIGISIGYVAHQHVSQFVLKEVEFRCHRTFARDYVVENKDNVISIDH
 RGAKSTALFASCYNTNIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFISSTNTSDSTCFLT
 RLNNAPFMDLNVSVYCEEHLHSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMGAFMDEIIGGT
 40 NKISQNTVLITDQSEEFDILRYSTLNTNAYDYFGKTLVYVLDPAFTTNRKASGTGVAAVGAYRHQFLIY
 GLEHFFLRDLSESSEVAIAECAAHMII SVLSLHPYLDELRIAVEGNTNQAAAVRIACLIRQSVQSSTLIR
 VLFYHTPDQNHIEQPFYLMGRDKALAVEQFISRFNSGYIKASQELVSYTIKLSHDPIEYLLLEQIQNLHRV
 TLAEGTTARYSAKRQNRISDDLIIAVIMATYLCDDIHAIRFRVS
 >gi_2746296
 45 MLRSCDIDAIQKAYQSIWKHEQDVKISSTFPNSAIFCQKRFIILTPELGFTTHAYCRHVKPLYLFCDRQR
 HVKSKIAICDPLNCALSKLKFTAIIEKNTEVQYQKHLELQTSFYRNPMLQIEKFIQDFQRWICGFENT

-15-

NKKERIKLEPFQKSILIHIIFFISVTKLPTLANHVLDYLKYKFDIEFINESSVNILKQKASVFLVPRRHG
 KTWFMIPVICFLLKNLEGISIGYVAHQKHVSFVMKDVEFKCRRFFPQKNITCQDNVITIEHETIKSTAL
 FASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQSSTKIIFISSTNSGNHSTSFLTCLNSNPF
 MLTVSVSYVCEDHVHILNDRGNATTACCYRLHKPKFISINADVKTADLFLEGAFKHEIMGGSCLCNVNDT
 5 LITEQGLIEFDLFRYSTISKQIIPFLGKELYIIDPAYTINRRASGTGVAAIGTYGDQYIIYGMEHYFLE
 SLLSNSDASIAECASHMILAVLELHPFFTELKIIIEGNSNQSSAVKIIACILKQTISVIRYKHITFFHTLD
 QSQIAQPFYLLGREKRLAVEYFISNFNSGYIKASQELISFTIKITYDPVEYALEQIRNIHQISVNN
 NAKKQTCSDDLLISIIIMAIYMCHEGKQTSFKEI
 >gi_325496
 10 MLRTCDITHIKNNEYAIIWKGERDCSTISTKYPNSAIFYKKRFIMLTPELGFAHSYNQVVKPLYTFCEKQ
 RHLKNRKLPTILPSLSHLQEMKFLPASDKSFESQYTEFLESFKILYREPLFLQIDGFIKDFRKWIKGEF
 NDFGDTRKIQLEPFQKNILIHVIFIAVTCLPALANRVINYLTHVFDIEFVNESTLNTLKQKTNVFLVPR
 RHGKTWFIPIISFLLKNIEGISIGYVAHQKHVSFVMKEVEFKCRRMFPEKTTITCLDNVITIDHQNKS
 TALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQASTKILFISSTNSGNHSTSFLMKLNN
 15 SPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKATANLFLEGAFIHEIMGGATCNV
 INDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIIVGMEH
 YFLESMTSSDTAIAECAAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSIQVTF
 HTPDQONQIAQPFYLLGKEKKLAVEFFISNFNSGNIKASQELISFTIKITYDPVEYALEQIRNIHQISVNN
 YITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 20 >gi_854039
 MKLNNSPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKATANLFLEGAFIHEIMGG
 ATCNVINDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIV
 YGMEHYFLESMTSSDTAIAECAAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSI
 QVTFHTPDQONQIAQPFYLLGKEKKLAVEFFISNFNSGNIKASQELISFTIKITYDPVEYALEQIRNIHQ
 25 ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_5733564
 MLRTCDITHIKNNEYAIIWKGERNCSTISTKYPNSAIFYKKRFIMLTPELGFAHSYNQVVKPLYTFCEKQ
 RHLKNRKLPTILPSLTRKLQEMKFLPASDKSFESQYTEFLESFKILYREPLFLQIDGFIKDFRKWIKGEF
 NDFGDTRKIQLEPFQKNILIHVIFIAVTCLPALANRVINYLTHVFDIEFVNESTLNTLKQKTNVFLVPR
 30 RHGKTWFIPIISFLLKNIEGISIGYVAHQKHVSFVMKEVEFKCRRMFPEKTTITCLDNVITIDHQNKS
 TALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQASTKILFISSTNSGNHSTSFLMKLNN
 PFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKATANLFLEGAFIHEIMGGATCNVI
 NDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIIVGMEHY
 FLESMTSSDTAIAECAAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSIQVTFH
 35 TPDQONQIAQPFYLLGKEKKLAVEFFISNFNSGNIKASQELISFTIKITYDPVEYALEQIRNIHQISVNNY
 ITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_4996048
 MKLNNSPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKATANLFLEGAFIHEIMGG
 ATCNVINDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIV
 40 YGMEHYFLESMTSSDTAIAECAAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSI
 QVTFHTPDQONQIAQPFYLLGKEKKLAVEFFISNFNSGNIKASQELISFTIKITYDPVEYALEQIRNIHQ
 ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_1136808
 MLLSRHRERLANLEETAKDAGERWELSAPTFTRHCPKTARMAHPFIGVVHRINSYSSVLETYCTRHHPA
 45 TPTSANPDVGTTPRPSEDNPAKPRLLLESSTYLQMRVCREDAHVSTADQLVEYQAGRKTHDSLHACS VYR
 ELQAFVLNLSFSLNGCYVPGVHWLEPFQQQLVMHTFFFLVSIKAPQKTHQLFGLFKQYFGLFETPNVSLQ

-16-

TFKQKASVFLIPRRHGKTWIVVAIIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLCRWFPPKNLNK
KENGTTIYTRPGGRSSSLMCATCFNKNSIRGQTFNLLYVDEANFIKKDALPAILGFMLQKDAKLIFISSV
NSSDRSTSFLNLNRNAQEKMLNVVSVCADHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
GAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQEVQKCLGKQLFVYIDPAYTNNTASGTGV
5 GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNVAAVEGNSSQDSGVA
IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQTVVSNITIKISFDPV
TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISPQ
>gi_1718281
MLQKDAKLIFISSVNSSDRSTSFLNLNRNAQEKMLNVVSVCADHREDFHLQDALVSCPCYRLHIPTYIT
10 IDESIKTTTNLFMEGAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYID
PAYTNNTASGTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNV
AAVEGNSSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQ
TVVSNITIKISFDPV TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISP
Q
15 >gi_2246515
MLQKDAKLIFISSVNSSDRSTSFLNLNRNAQEKMLNVVSVCADHREDFHLQDALVSCPCYRLHIPTYIT
IDESIKTTTNLFMEGAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYID
PAYTNNTASGTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNV
AAVEGNSSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQ
20 TVVSNITIKISFDPV TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISP
Q
>gi_2246552
MLLSRHRERLAANLQETAKDAGERWELSAPTFRHCPKTARMAHPFIGVVHRINSYSSVLETYCTRHPA
TPTSANPDVGTTPRSEDNVPKPRLLESSTYLQMCVREDAHVSTADQLVEYQAARKTHDSLHACSVYR
25 ELQAFVLNLSFNGCYVPGVHWLEPFQQQLVMHTFFFLVSIKAPQKTHQLFGLFKQYFGLFETPNSVLQ
TFKQKASVFLIPRRHGKTWIVVAIIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLCRWFPPKNLNK
KENGTTIYTRPGGRSSSLMCATCFNKNSIRGQTFNLLYVDEANFIKKDALPAILGFMLQKDAKLIFISSV
NSSDRSTSFLNLNRNAQEKMLNVVSVCADHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
GAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYIDPAYTNNTASGTGV
30 GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNVAAVEGNSSQDSGVA
IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQTVVSNITIKISFDPV
TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISPQ
>gi_4494933
MLQKDAKLIFISSNSSDKSTSFLNLNKAHEKMLNVVNYVCPDHKDDFNLDQTVVACPCYRLHIPAYIT
35 IDETVRSTTNLFLEGAFTSTELMGDAATSAQSMHKIVSDSSLSQLDLCRVKSTSQDIQGAMKPCPLHVYIDP
AYTNNTDASGTGIGAVIAVNHKVIKICILLGVEHFFLRDLTGTAAYQIASCAAALIRAI VTLHPQITHVNV
AVEGNSSQDAGVAIATVLNEICSVPLSFLHHVDKNTLIRSPIYMLGPEKAKAFESFIYALNSGTFSASQT
VVSHTIKLSFDPVAYLIDQIKAIRCIPLKDGHTYCAKQKMSDDVLVAAVMAHYMATNDKFVFKSLE
>gi_7330018
40 MLQKDAKLIFISSNSSDKSTSFLNLNKAHEKMLNVVNYVCPDHKDDFNLDQTVVACPCYRLHIPAYIT
IDETVRSTTNLFLEGAFTSTELMGDAATSAQSMHKIVSDSSLSQLDLCRVESTSQDIQGAMKPCPLHVYIDP
AYTNNTDASGTGIGAVIAVNHKVIKICILLGVEHFFLRDLTGTAAYQIASCAAALIRAI VTLHPQITHVNV
AVEGNSSQDAGVAIATVLNEICSVPLSFLHHADKNTLIRSPIYMLGPEKAKAFESFIYALNSGTFSASQT
VVSHTIKLSFDPVAYLIDQIKAIRCIPLKDGHTYCAKQKMSDDVLVAAVMAHYMATNDKFVFKSLE
45 >gi_4019255
MLLLKAKKALMENL TEASSTQSETEWTVDTPTMITNIKKSERMAYSKIGVIPSINLYSASLTSFCRLYRP

-17-

LALKQPLPQTGTLRLLPSEKPYISQKLSNYVKSLLTLKHVWHDIEAEAEYYASVQTEKTFMECPYIYLELRQ
FIINLSSFLNGCYVKKSTHIEPFQLQLILHTFYFLISIKSPESTNKLFDIFKEYFGLGEMDSAMLQNFQKQ
KASIFLIPIRRHGKTWIVVAIISMLITSVENLHVGYVAHQKHVANSVFTEIINTLQKWFPSPKNIDVKKENG
TIIYKIPGKKPSTLMCASC FNKNSIRGQTFNLLYIDEANFIKKDSLPAILGFMLQKDAKLIFISSVNSGD
5 KATSFLFNLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYITIDETIKNTTNLFLDGAFT
TELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMYLYIDPAYTNNSEASGTGIGAI
ALKNNSSKCIIVGIEHYFLKDLTGATYQIASCACSLIRAALVLYPHIQAVHVAVEGNSSQDSAVAISTF
LNECSPVKVNFMHYKDKTTAMQWPIYMLGSEKSAFESFIYAINSGTISASQSIISNTIKLTFDPISYLI
EQIRAIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAHFFSTSNKHIFKQLNSI
10 >gi_4019257
MLQKDAKLIFISSVNSGDKATSFLFNLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYIT
IDETIKNTTNLFLDGAFTTELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMYLYID
PAYTNNSEASGTGIGAIALKNNSSKCIIVGIEHYFLKDLTGATYQIASCACSLIRAALVLYPHIQAVH
VAVEGNSSQDSAVAISTFLENCSPVKVNFMHYKDKTTAMQWPIYMLGSEKSAFESFIYAINSGTISASQ
15 SIISNTIKLTFDPISYLIEQIRAIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAHFFSTSNKHIFKQLNS
I
>gi_60355
MLLLKAKKAIENLSEVSSTQAETDWMSTPTIITNTSKSERTAYSIGVIPSUNLYSSTLTSFCKLYHP
LTLNQTQPTGTLRLLPHEKPLILQDLSNYVKLLTSQNVCHDTEANTEYNAAVQTQKTSMECPTYLELRQ
20 FVINLSSFLNGCYVKKSTHIEPFQLQLILHTFYFLISIKSPESTNRLFDIFKEYFGLREMDPDLQIFKQ
KASIFLIPIRRHGKTWIVVAIISMLLTSVENIHVGYVAHQKHVANSVFTEIINTLQKWFPSPRYIDIKKENG
TIIYKSPDKKPSTLMCATCFNKNSIRGQTFNLLYIDEANFIKKDSLPAILGFMLQKDAKLIFISSVNSGD
RATSFLFNLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYITIDETIKNTTNLFLDGAFT
TELMGDMGSGISKSNMHKVISEMAITQFDLCRADTTKPEITQCLNSTMYIYIDPAYTNNSEASGTGIGAIL
25 TFKNNSSKCIIVGMEHYFLKDLTGATYQIASCACSLIRASLVLYPHIQCVHVAVEGNSSQDSAVAISTL
INECSPKIVFYIHYKDKTTMOWPIYMLGAESIAFESFIYAINSGTISASQSIISNTIKLSFDPISYLI
EQIRSIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAYFFATSNKHIFKPLNST
>gi_695201
MLQKDAKIIIFISSVNSSDQTTSFLYNLKNASEKMLNVVNYVCPQHREDFSLQESVVS CPCYRLHIPTYIA
30 IDENIKDTTNLFMEGAFTTELMGDGAAATTQTNMHKVVGEPALVQFDLCRVDTGSPEAQRGLNPTLFLYV
DPAYTNNTEASGTGMGAVVSMKNSDRCVVGVGEHFFLKELTGASSLQIASCAAALIRSLATLHPFVREAH
VAIEGNSSQDSAVAIAITLLHRSPLPVKFLHHDKATGVQWPMYILGAEKARAFETFIYALNSNTLS CGQ
AIVSNTIKLSFDPVAYLIEQIRAIKCYPLKDGTVSYCAKHKGSDDTLVAVVMAHYFATSDRHVFNHMK
QI
35 >gi_4928934
MLLSSFRNLQKNYKYSVQAQONIDWPVETPVLISKDSKTNRLAHPLIGVISRINLYSPTLKYYCDEYST
TKQPKFTPDIGYVRDLKKHDQYFLPKLQHLSTLCEAYNHVDRQAQVEFNASILTAKAFNANGVLNELKQ
FLINLSCFLNGCYVSKSTCIELFQQLILHTFYFLISIKTPEETNKMFTFFKHVGLFDIDDNMLQCFKQ
KSTVFLIPIRRHGKTWIVVAIISVLLASVENVHIGYVAHQKHVANAVFTEIITTLYQWFPSPKNIEIKKENG
40 TIIYTKPGRKPSTLMCATCFNKNSIRGQTFNILYVDEANFIKKEALPAILGFMLQKDAKIIIFISSVNSAD
KSTSFLFNLNAKEKMLNVVNYVCPHKKEDFNLSSTLSCPCYRLHIPTYITIDESIKNTTNLFLDDVFT
TELMGDISFTPTSSMFKVVEEQALFHFDICRVDTTQIDTVKIIDNVLYVYVDRAYTSNSEASGTGIGAVV
PLKTKVKTIILGIEHYFLKNLTGTASQQIAYCVTSMIKAILTLHPHINHVNVAVEGNSSQDSAVAISTFI
NEYCPVPVFFAHCNERSVFWQWPIYILGSEKSAFEKFCALNTGTLSASQTIVSNTIKISFDPVAYLME
45 QIRAIRCLPLKDGSYTYCAKQKMSDDTLVAVVMANYMAISEKHTFKELCKT
>gi_1632798

-18-

MLYASQGRGLTENLRNALQQDSTTQGCGLGAETPSIMYTGAKSDRWAHPLVGTIHASNLYCPMLRAYCRHY
 GPRPVFVASDESLEPMFGASPALHTPVQVQMCLLPCLRLDRLQRLPPNLEDSEALTEFKTSVSSARAILE
 DPNFLEMREFVTSLASFLSGQYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMT
 LEKLHIFKQKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKSFDSK
 5 RVEVNKETSTITFRHSGKISSTVMCATCFNKNSIRGQTFHLLFVDEANFIKKEALPAILGFMLQKDAKII
 FISSVNSADQATSFLYKLKDAQERLLNVVSVCQEHQDFDMQDSMVSCPCFRLHIPSYITMDSNIRATT
 NLFLDGAFTSTELMGDTSSLSQGSLSRTVRDDAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSA
 SGTGIAAVTHDRADPNRVIIVLGLHFFLKDLTGDAALQIATCVVALVSSIIVTLHPHLEEVKVAVEGNSSQ
 DSAVAIASIIGESCPLPCAFVHTKDKTSSLQWPMYLLTNEKSKAERLIYAVNTASLSASQVTVSNTIQL
 10 SFDPVLYLISQIRAIKPIPLRDGTYTYTGKQKNLSDDVLVALVMAHFLATTQKHTFKKVH
 >gi_2337991
 MFYVKVMPALQKACEELQNWSAKSGKWPVPETPLVAVETRRESERWPHPYLGLLPGVAAYSSTLEDYCHL
 YNPYIDALTRCDLGQTHRRVATQPVLSDDQLCQQLKFLSCPRNTSVKAKLEFEAAVRTHQALDNSQVFLE
 LKTFVLNLSAFLNKRYSRSSHIEFLQKQLIMHTFFFLVSIKAPELCEKFCNIFKLYFNIDTMDQATLDI
 15 FKQKASVFLIPRRHGKTWIVVAIISILLASVQDLRIGYVAHQKHVANAVFTEVINLHTFFPGKYMDVKK
 ENGTIIFGLPNKKPSTLLCATCFNKNSIRGQTFQLLFVDEANFIKDALEPTILGFMLQKDAKIIFISSSN
 SSDQSTSFLYNLKGASERMLNVVSVCNHNKEDFSMQDGLISPCYSLHVPYSYISIDEQIKTTTNLFLDG
 VFDTELMGDSGCGTLSTFQIISSEALSQFELCRIDTASPOVQAHNLSTVHMYIDPAFTNNLDASGTGISV
 IGRLGAKTKVILGCEHFFLQKLTGTAAALQIASCATSLLRSVVIHPMIKCAQITIEGNSSQDSAVAIANF
 20 IDECAPIPVTFYHQSDKTKGVLCPLVLLGQEKAVAFESFIYAMNLGLCKASQLIVSHTIKLSFDPVTYLL
 EQVRAIKCQSLRDGSHTYHAKQKNLSDDLVSVMMSLYLSSANTLPFKPLHIERFF
 >gi_2317977
 MLQKDAKIIFFISSVNSGEKTTSTFLYNLKDANEKMNVSVCSEHMEDEFNKQSAITACPCYRLVPEFIT
 INDNIKCTTNLLLEGSFATELMGNMQSHTEVSGNSMIHESSLTRLDIFYRCDTAGQGAPTENTLFFVYIDP
 25 AYGNNVHASGTGIVAMSHCKHTKKCIILGLEHFFLNLTGTAAHNIASCATALLEGILFQHPWIQEIRCI
 IEGNSNQDSAVAIATFISHNIKLPFLFASYRDKTGMQWPIYMLSGDKTLAFQNFISSLNQGLLCASQTVV
 SNTVLLSSDPISYLLIEQIKNTKCIYHKNKTITTFQSKTHTMSDDVLIACVMTCYVMTTNKISYISFSIK
 >gi_6625593
 MFIASKKSYFEAVRSTVSSHSEEFWKSDDPVYFTQYKKQCNRLPNAYLGTLSASKYSENFRHYVATFS
 30 NSPLDFPQSVFNERNPCEYSVPYLDALQCSAKTLVGCSVSTTERNEYEVCKEATRCFKDAMSHKVLKVF
 LSNLSWFLKGHYKSKQAFLEPFQKQLILHSFMFVASIKCPETTTKLDFEFKFLDMLYFDNTDLLTFLQK
 SPAFLIPRRHGKTWIVTAIISMLLTSVDDLHIGYVAHQKHVSLAVFLEISNILLAWFPRKNIDIKKENG
 ILYSHPGKKSSTLMCATCFNKNSIRGQTFNLLFVDEANFIKKEALPAILGFMLQKDAKIIFFISSVNSGEK
 TTSFLYNLKDANEKMNVSVCSEHMEDEFNKQSAITACPCYRLVPEFITINDNIKCTTNLLLEGSFAT
 35 ELMGNMQSHTEVSGNSMIHESSLTRLDIFYRCDTAGQGAPTENTLFFVYIDPAYGNNVHASGTGIVAMSHC
 KHTKKCIILGLEHFFLNLTGTAAHNIASCATALLEGILFQHPWIQEIRCIIEGNSNQDSAVAIATFISH
 NIKLPFLFASYRDKTGMQWPIYMLSGDKTLAFQNFISSLNQGLLCASQTVVSNVLLSSDPISYLLIEQIK
 NTKCIYHKNKTITTFQSKTHTMSDDVLIACVMTCYVMTTNKISYISFSIK

gi_10180719	MFGLGEEETKRHFERLHKTNDRLGASHNRKSIROG---	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515
-------------	--	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

gi_10180719	131	RITNAQFLSPHMHRYVANSIFQYOEHECAHHRHAYTSTHNSFISHRISDAFKQLTVFISRFSKLLTASFRDYNKLDHHTYK---KRARIDAPSYOKLHGTLLEFQKHTLHARTYFVTSVLLGD-HAERA	260
gi_7673189	140	RTTNAQFLSPHMHRYVANSIFQYOEHECAHHRHAYTSTHNSFISHRISDAFKQLTVFISRFSKLLTASFRDYNKLDHHTYK---KRARIDAPSYOKLHGTLLEFQKHTLHARTYFVTSVLLGD-HAERA	250
gi_5689285	150	SKNRREALLAPTLTRYANSITFHEYUQAQCAHHRHAYTSTHNSFISHRISDAFKQLTVFISRFSKLLTASFRDYNKLDHHTYK---KRARITAPSYOKPHGTLLEFQKHTLHARTYFVTSVLLGD-HAERA	240
gi_1869837	160	--TAKLEFLAPELVRAVARLRFRECAPEDVPPQRHAYYSVLNTHQALHSEARQLVHFYRDFRQLLTSFRASSLAEITGPP--KKRAKVDVATHTGRTYGTLEFQKHTLHARTYFVTSVLLGD-HAERA	230
gi_59501	170	--TAKLEFLAPELVRAVARLRFRECAPEDVPPQRHAYYSVLNTHQALHSEARQLVHFYRDFRQLLTSFRASSLAEITGPP--KKRAKVDVATHTGRTYGTLEFQKHTLHARTYFVTSVLLGD-HAERA	220
gi_2605992	180	--SSVEARYVDPKVKQALKTLSFVEYNDEARSCRNAYYSVLNTHQALHSEARQLVHFYRDFRQLLTSFRASSLAEITGPP--KKRAKVDVATHTGRTYGTLEFQKHTLHARTYFVTSVLLGD-HAERA	210
gi_330792	190	--ASVEARYVDPKVKQALKTLSFVEYNDEARSCRNAYYSVLNTHQALHSEARQLVHFYRDFRQLLTSFRASSLAEITGPP--KKRAKVDVATHTGRTYGTLEFQKHTLHARTYFVTSVLLGD-HAERA	200
gi_971317	200	--GTGEARHVSRELAVLSALRFARHAPPAEAAHCHAYSYHAALSHRASGAFAGVHAFVAREFSLVDTSFNGADLDGQGQAT--SKRKYVDVPTYGKQRTLEFQKHTLHARTYFVTSVLLGD-HAERA	190
gi_5869808	210	--YSVHAHVSPELCKHAYSSVQFYEYSPPEERAPHRHAYSGVHHTFRASLSDSFCQLSTFTQRFSYLVEISFESIECCSGHG---KRAKVDVPTYGKQRTLEFQKHTLHARTYFVTSVLLGD-HAERA	180
gi_5708110	220	--FTGAEDSTLPAIROKLAHLNFAFAPSDSLHDKAFDGINHGYRBFYKSDQFSQNNFTYRFTHTLLKKSFGQASHDY---KRAKLEKTTSEQRDGTLEFQKHTLHARTYFVTSVLLGD-HAERA	170
gi_1013970	230	-----LFHERKLSALDKLTFRQCSEEQ-R---ASYOK-LDALTELYRQDQFQQINNFTDQKALDGGFSTAVEGD---KRAKLEKTTSEQRDGTLEFQKHTLHARTYFVTSVLLGD-HAERA	160
gi_2746296	240	-----RICOPLNICALSKLKTALIEKNT-----VQYQKHALELQTSFYRHPMFLQIEKFTQDFQRHICGQFENT--HK---KRAKLEKTTSEQRDGTLEFQKHTLHARTYFVTSVLLGD-HAERA	150
gi_325496	250	-----TILPSLTKLQEHKFLPASDKSFE---SQYTEFLESFKILYREPLFLQIDGFIKDFKXIKGEEFNDF--GD---KRAKLEKTTSEQRDGTLEFQKHTLHARTYFVTSVLLGD-HAERA	140
gi_5733564	260	-----TILPSLTKLQEHKFLPASDKSFE---SQYTEFLESFKILYREPLFLQIDGFIKDFKXIKGEEFNDF--GD---KRAKLEKTTSEQRDGTLEFQKHTLHARTYFVTSVLLGD-HAERA	130
gi_1136808	270	DNVPAKPRLLLESSTYLQHRCPREDHAYSTADQLVEYQAGRKTHDSLHACSVYRELQAFVNLSSFLNGCYVP---GVHLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	120
gi_2246552	280	SE---KPYISQKLSHYVKSLLTQKAVHDIIEH---AEYASVQTEKTFHECPTYLELRQFVNLSSFLNGCYVP---GVHLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	110
gi_4019255	290	HE---KPYISQKLSHYVKSLLTQKAVHDIIEH---AEYASVQTEKTFHECPTYLELRQFVNLSSFLNGCYVP---GVHLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	100
gi_60355	300	KH---DQYFLPKLQHMLSTLCERYHNVDRQQA---VEFNASTLTKAFNANGVNLKQLFVNLSSFLNGCYVP---GVHLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	90
gi_4928934	310	AT---DQYFLPKLQHMLSTLCERYHNVDRQQA---VEFNASTLTKAFNANGVNLKQLFVNLSSFLNGCYVP---GVHLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	80
gi_2337991	320	TPVQVQNCLLPELRDTLQRLLPPNLEDSEAL---TEFKTSVSSARAILEDPINFLEHREFVTSLSASFLSGQYKH---KQAFLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	70
gi_1632798	330	YSV---PYLDSALQCSAKTLVGCYSYSTIER---NEVEYCKEATKCFKDMHSHKVLKVFLSNLSHFLKGHYKS---KQAFLEPFEQQLVHHTFFELVSTKAPQ-KTHQL	60
gi_6625593	340		50
gi_171281	350		40
gi_2246515	360		30
gi_4494933	370		20
gi_7330018	380		10
gi_4019257	390		0
gi_695201	400		
gi_2317977	410		
gi_854039	420		
gi_4996048	430		

TABLE 2 CONTINUED

gi_101801719	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	261
gi_7673189	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	270
gi_5689285	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	280
gi_1869837	HTFLRLVFEIPLFSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	290
gi_59501	HTFLRLVFEIPLFSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	300
gi_2605992	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	310
gi_330792	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	320
gi_971317	DCFLRTVFNTPSYSDVLEHFKQKSTVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	330
gi_5869808	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	340
gi_5708110	NRYLSTVFNTPSYSDVLEHFKQKSTVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	350
gi_1813970	LOYLTHAFQIDFELSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	360
gi_2746296	LOYLTHAFQIDFELSDIVTRHFQRATVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	370
gi_325496	INYLTHVFDIEFVNESTLNLKQKTNVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	380
gi_5733564	INYLTHVFDIEFVNESTLNLKQKTNVFLVPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	390
gi_1136808	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_2246552	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_4013255	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_60355	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_4928934	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_2337991	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_1632798	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_5625593	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_1718281	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_2246515	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_4494933	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_7330018	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_4013257	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_695201	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_2317977	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_854039	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
gi_4996048	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	
Consensus	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTUFLVPLIALAHSSFEIGIRIGYTSHTKRIEVPFEDIGURLRHFGRHRYDVHVKGE-TITFSFPGSLKSTVTFASSHNTN-SIRGQDFNLLFVD	

TABLE 2 CONTINUED

[illegible]

TABLE 2 CONTINUED

gi_10180719	DRFLYRPSTVNNQDP--FSQDLVYVDPAPFANTKRSCTGAVVIGKYGTD--YIVFGLHLYFLRALTGESSDSIGYVACQCLIQICATIRKRFQVTKIATEGNSQDSAVAIATLRIATISYHKRAV	521	530	540	550	560	570	580	590	600	610	620	630	640	650
gi_7673189	DRFLYRPSTVNNQDP--FSQDLVYVDPAPFANTKRSCTGAVVIGKYGTD--YIVFGLHLYFLRALTGESSDSIGYVACQCLIQICATIRKRFQVTKIATEGNSQDSAVAIATLRIATISYHKRAV														
gi_5689285	ERLLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_1869837	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_59501	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2605992	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_330792	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_971317	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_5869808	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_5708110	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_1813970	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2746296	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_325496	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_5733564	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_1136808	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2246552	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_4019255	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_60355	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_4928934	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2337991	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_1632798	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_6625593	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_1718281	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2246515	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_4494933	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_7330018	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_4019257	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_695201	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_2317977	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_854039	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
gi_4996048	ERFLLYRPSTVNNQDI--LSQDLVYVDPAPFANTKRSCTGAVVIGRYGAD--YITFGLHLYFLRALTGESAPRIAGCAQCTAQCACNCEPFGITIRVAVEGNSQDSAVAIATLRIATISYHKRAV														
Consensus	.qfd.cr.de.....1...\$.vYIDP4ZnH..#SGT61.av.....ii.g.EHZEI..ltg.aa...i.i..lhp..i.v..a!EGNSqdsav!at.....e.....														

TABLE 2 CONTINUED

[illegible]

TABLE 2 CONTINUED

```

781      791
|-----+|
gi_10180719
gi_7673189
gi_5689285
gi_1869837 TS
gi_59501 VS
gi_2605992 q
gi_330792 q
gi_971317 RPPAD
gi_5869808
gi_5708110 IGRKQRP
gi_1813970
gi_2746296
gi_325496
gi_5733564
gi_1136808
gi_2246552
gi_4019255
gi_60355
gi_4928934
gi_2337991 FF
gi_1632798
gi_6625593
gi_1718281
gi_2246515
gi_4494933
gi_7330018
gi_4019257
gi_695201
gi_2317977
gi_854039
gi_4986048
Consensus ..

```

TABLE 2 CONTINUED

-26-

Table 3. Degenerate primers generated by CODEHOP

Block x7263xbliD

T L Y V Y I D P
 oligo:5'-AACCTGTACGTGtayntngaycc-3' degen=64 temp=33.4 Extend clamp

T L Y V Y I D P A
 oligo:5'-AACCTGTACGTGTACntngayccngc-3' degen=128 temp=36.0 Extend clamp

T L Y V Y I D P A Y
 oligo:5'-AACCTGTACGTGTACATngayccngcnt-3' degen=128 temp=42.5 Extend clamp

Complement of Block x7263xbliD

Y I D P A Y T N N T
 atrnanctrggGCGGATGTGGTTGTGT oligo:5'-TGTGTGTGGTGTAGGCGgggrtcnanrta-3'
 degen=64 temp=62.9

D P A Y T N N T R A
 anctrngcngnaTGTGGTTGTGTGGGTCCG oligo:5'-GCCTGGGTGTGTGTGGTGTangcnggrtcna-3'
 degen=128 temp=61.8

D P A Y T N N T R A
 ctrngcngnawGTGGTTGTGTGGGTCCG oligo:5'-GCCTGGGTGTGTGTGGTGWangcnggrtc-3'
 degen=64 temp=61.0

Block x7263xbliE

C I I F G M E H F F
 oligo:5'-TGGATCATCTTCGGCATngarcaytwyt-3' degen=64 temp=55.7 Extend clamp

I F G M E H F F L
 oligo:5'-CATCTTCGGCATGGAGcaytwytwyyt-3' degen=64 temp=62.0

Complement of Block x7263xbliE

E H F F L R D L T G
 ctygtrawrawGGACTTCCTGGACTGCCC oligo:5'-CCCGTCAGGTCCTTCAGGwarwartgytc-3'
 degen=32 temp=61.7

H F F L R D L T G
 tygtrawrawrrrACTTCCTGGACTGCCCC oligo:5'-GCCCCTCAGGTCCTTCarrwarwartgyt-3'
 degen=128 temp=60.8

H F F L R D L T G
 gtrawrawrrraCTTCCTGGACTGCCCC oligo:5'-GCCCCTCAGGTCCTTCarrwarwartg-3' degen=64
 temp=60.8

Block x7263xbliF

E V H I A V E G N
 oligo:5'-GGACGTGCACGTCGCCrtnrgarggnaa-3' degen=64 temp=63.8

Complement of Block x7263xbliF

E G N S S Q D S A
 anctyccnttrwGGTTGGTCCTGAGGCGG oligo:5'-GGCGGAGTCCTGGTTGGwrttnccytcna-3'
 degen=128 temp=62.7

E G N S S Q D S A V
 ctyccnttrwsGTTGGTCCTGAGGCGGC oligo:5'-CGGCGGAGTCCTGGTTGswrttnccytc-3'
 degen=64 temp=63.9

-27-

CLAIMS

1. A method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR,
5 which method comprises

(a) providing a plurality of amino acid sequences from members of a first virus family,

10

(b) comparing the sequences to identify conserved regions,

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region identified in

15 (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

20

2. A method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR,
25 which method comprises

(a) providing a plurality of nucleotide sequences from members of a first virus family,

30 (b) comparing the sequences to identify conserved regions,

-28-

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a conserved region identified in (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the conserved regions in members of the first virus family, and

5

(d) repeating steps (a) to (c) for each virus family.

3. A method according to claim 1 or 2 which further comprises synthesising one or more of the primer pairs and determining optimal conditions for using the primer

10 pairs in PCR.

4. A method according to any one of the preceding claims which comprises testing the ability of one or more of the primer pairs to amplify a nucleotide sequence that encodes an amino acid as defined in claim 1(a) or a nucleotide sequence as defined in

15 claim 2(a).

5. A method according to claim 3 or 4 which comprises testing the primer pair(s) in a range of buffer conditions to determine the optimal buffer conditions for PCR.

20 6. A method according to any one of claims 3 to 5 which comprises testing the primer pair(s) at a range of different temperatures to determine the optimal temperature for PCR.

7. A method according to any one of the preceding claims which comprises

25 identifying one or more groups of primer pairs wherein the primer pairs in each group have similar optimal conditions of use in PCR such that they can be used optimally in the same reaction vessel.

8. A method according to claim 7 wherein each primer pair in a group generates a

30 PCR product of a different size to the other primer pair(s) in the group.

-29-

9. A method according to claim 7 or 8 wherein each primer pair in a group carries a different label from the other primer pair(s) in the group.
10. A method according to claim 9 wherein each primer pair in a group carries a
5 differently-coloured fluorescent label.
11. A method according to any one of the preceding claims wherein the biological sample is a single-source sample from a single individual or is a pooled sample from more than one individual of the same species.
- 10 12. A method according to claim 11 wherein the biological sample is a human sample.
13. A method according to any one of the preceding claims wherein at least 50% of
15 the primer pairs bind a sequence that is conserved across all of the genres and/or subfamilies.
14. A panel of primers designed according to any one of the preceding claims.

FIG 1

1/1

